

Input file F1h14273new; Duput File F1h14273tra  
Sequence length 1743

TCCGGACTAGTTCTAGACCGCTGCGGGCCGCGCCAGCGCGGGGA	ATG	TCC	CCT	GAA	TGC	C	A	R	A	A	GCG	9
G D A P L R S L E Q A N R T R F F S												27
GGC GAC GCG CCC TTG CCG AGC CTG GAG CAA GCC AAC CGC ACC CGC TTT CCC TTC TTC TCC												29
D V K G D H R L V L A A V E T T V L V L												87
GAC GTC AAG GGC GAC CAC CCG CTG GTG CTG GCG GCG GTG GAG ACA ACC GTG CTG GTG CTC												49
I F A V S L L G N V C A L V L V A R R R												147
ATC TTT GCA GTG TCG CTG GCG AAC GTG TGC GCG CTG GTG CTG GTG GCG CGC CGA CGA												69
R R G A T A C L V L N L F C A D L L F I												207
CGC CGC GGC GCG ACT GCC TGC CTG GTA CTC AAC CTC TTC TGC GCG GAC CTG CTC TTC ATC												89
S A I P L V L A V R W T E A W L L G P V												267
AGC GCT ATC CCT CTG GTG CTG GCG GTG GCG TGG ACT GAG GCC TGG CTG CTG GCG CCC GTT												109
A C H L L F Y V M T L S G S V T I L T L												327
GCC TGC CAC CTG CTC TTC TAC GTG ATG ACC CTG AGC GGC AGC GTC ACC ATC CTC ACG CTG												129
A A V S L E R M V C I V H L Q R G V R G												387
GCC GCG GTC AGC CTG GAG CCG ATG GTG TGC ATC GRG CAC CTG CAG CCG GGC GTG CCG GGT												149
P G R R A R A V L L A L I W G Y S A V A												447
CCT CCG CCG GCG GCG GCA GTG CTG CTG GCG CTC ATC TGG GCC TAT TCG GCG GTC GCG												169
A L P L C V F F R V V P Q R L P G A D Q												507
GCT CTG CCT CTC TGC TTT TTT CGA GTC GTC CCG CAA CCG CTC CCC GCG GCG GAC CAG												567

TO FIG. 1B.

**FIG. 1A.**





10075987 .072202

2/15

COPY OF PAPERS  
ORIGINALLY FILED

FROM FIG. 1A.

E	I	S	I	C	T	L	I	W	P	T	I	P	G	E	I	S	W	D	V	209
GAA	ATT	TCG	ATT	TGC	ACA	CTG	ATT	TGG	CCC	ACC	ATT	CCT	GGA	GAG	ATC	TCG	TGG	GAT	GTC	627
S	F	V	T	L	N	F	L	V	P	G	L	V	I	V	I	S	Y	S	K	229
TCT	TTT	GTT	ACT	TTG	AAC	TTC	TTG	GTG	CCA	GGA	CTG	GTC	ATT	GTG	ATC	AGT	TAC	TCC	AAA	687
I	L	Q	I	T	K	A	S	R	K	R	L	T	V	S	L	A	Y	S	E	249
ATT	TTA	CAG	ATC	ACA	AAG	GCA	TCA	AGG	AAG	AGG	CTC	ACC	GTA	AGC	CTG	GCC	TAC	TCG	GAG	747
S	H	Q	I	R	V	S	Q	Q	D	F	R	L	F	R	T	L	F	L	L	269
AGC	CAC	CAG	ATC	CGC	GTG	TCC	CAG	CAG	GAC	TTC	CGG	CTC	TTC	CGC	ACC	CTC	TTC	CTC	CTC	807
M	V	S	F	F	I	M	W	S	P	I	I	I	T	I	L	L	I	L	I	208
ATG	GTC	TCC	TTC	TTC	ATC	ATG	TGG	AGC	CCC	ATC	ATC	ATC	ACC	ATC	CTC	CTC	ATC	CTG	ATC	867
Q	N	F	K	Q	D	L	V	I	W	P	S	L	F	F	W	V	V	A	F	309
CAG	AAC	TTC	AAG	CAA	GAC	CTG	GTG	ATC	TGG	CCG	TCC	CTC	TTC	TTC	TGG	GTG	GTG	GCC	TTC	927

TO FIG. 1C.

FIG. 1B.



10075987.072202

3/15

COPY OF PAPERS  
ORIGINALLY FILED

FROM FIG. 1B.

T F A N S A L N P I L Y N M T L C R N E 329  
ACA TTT GCT AAT TCA GCC CTA AAC CCC ATC CTC TAC AAC ATG ACA CTG TGC AGG AAT GAG 987  
W K K I F C C F W F P E K G A I L T D T 349  
TGG AAG AAA ATT TTT TGC TGC TTC TGG TTC CCA GAA AAG GGA GCC ATT TTA ACA GAC ACA ACA 1047  
S V K R N D L S I I S G . 362  
TCT GTC AAA AGA AAT GAC TTG TCG ATT ATT TCT GGC TAA 1086

TTTTTCTTTATAGCCGAGTTTCTCACACCTGGCGAGCTGTGGCATGCTTTTAAACAGAGTTCATTTCAGTACCCTCCA  
TCAGTGCACCCTGCTTTAAGAAAAAGAACCTATGCAATAGACATCCACAGCGTCGGTAAATTAAAGGGTGATCACCAA  
GTTTCATAATTTTCCCTTTATAAAGGATTGTGGCCAGGTGCAGTGGTTCATGCCTGTAATCCAGCAGTTTGGG  
AGGCTGAGGTGGTGATCACCTGAGGTCAGGAGTTCGAGACCAACCTGACCAACATGGTGAGACCCCGTCTCTACTA  
AAATAAAAAAAATAGCTGGGAGTGGTGGGACCTGTATCCTAGCTACTTGGGAGGCTGAACCCAGGAGAAAT  
CTCTTGAACTGGGAGGCAGAGGTTCAGTGAGCCGAGATCGTGCCATTGCACCTCCAAACAGGGCAACAAGAGTGAAAC  
TCCATCTTAAAAAAAGAGATTGTATGGGTCCCTTTAAATGTGAACTTTTTAGTGTTGTGTAATATG  
ATCAAAATTTAATAATATTTATTTATGACTGTTTCAGCAAAAAAAAGGGGGG

RTA01/2057957v1

FIG. 1C.



10075987.072202

4/15

COPY OF PAPERS  
ORIGINALLY FILED

Sequence	Description	score	E-value	N
7tm_1	PF00001 7 transmembrane receptor (rhodopsin	119.9	4.7e-37	1

Parsed for domains:

Sequence	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value	-
7tm_1	1/1	57	321	1	259 [ ]	119.9	4.7e-37	

Alignments of top-scoring domains:

7tm\_1: domain 1 of 1, from 57 to 321: score 119.9, E = 4.7e-37

\*->GN1LVilvi1rtkk1rtptnifi1NLAVADLLf11t1ppwalyylvg  
GN+ +++++tr +tr +t +++1NL ADLLf + p++ ++ -+

F1h14273, 57 GNVCALVLVAR-RRRRGATACLVNLF CADLLFISAIPVLAVR-WT 101

gaadWpfGsa1Ck1vtaldvvnmyaSi11Lta1SiDRY1A1vhP1ryrrr  
e W++G++ C+1+ ++++++ + 11+L+a S++R + Iv 1+ +r

F1h14373, 102 --EAWLLGPVACHLLFYMTLSGSVTILTAAVSLERMVCIV-HLQRGVR 148

rtsprrrAkvvil1vwvial11s1Pp11fswvktveegngt1nnvnevC1i  
+r +v+++1+W +++++1P +f+ v+ ++ ++ ++ +C++

F1h14273, 149 GPGRRARAVLLALIWGYSAAALPLCVFFRVVPQRLPG--ADQEISICTL 196

dfpccstasvstwlrsvy11st1wgF11P11vilvcYtr11rt1r....  
+p++++ ++s+ +++ ++ F1+P 1vi++ Y+ I1 + + +++++

F1h14273, 197 IWPTIPG--EISWDVSFVTLNGLVRLVIVISYSKILQITKasrkr 240

.....kaakt11vvvvvFv1CW1Pyfiv111dt1c  
+ + +++++ + +++++ ++ +t1++++v F++ W P i++11 +

F1h14273, 241 1cvslayseehqirvsqddFRLFRTLFLLMVSFFIMWSP11ITILLILIQ 290

.1si1msstCelerv1ptailvt1wLayvNsc1NPi1Y<-\*  
-+ + + p +++++ + +++++Na+1NPi+Y

F1h14273, 291 nFK-----QDLVIWPSLFFWVVPATFANSALNPILY 321

FIG. 2.



10075987 .072202

COPY OF PAPERS  
ORIGINALLY FILED

5/15

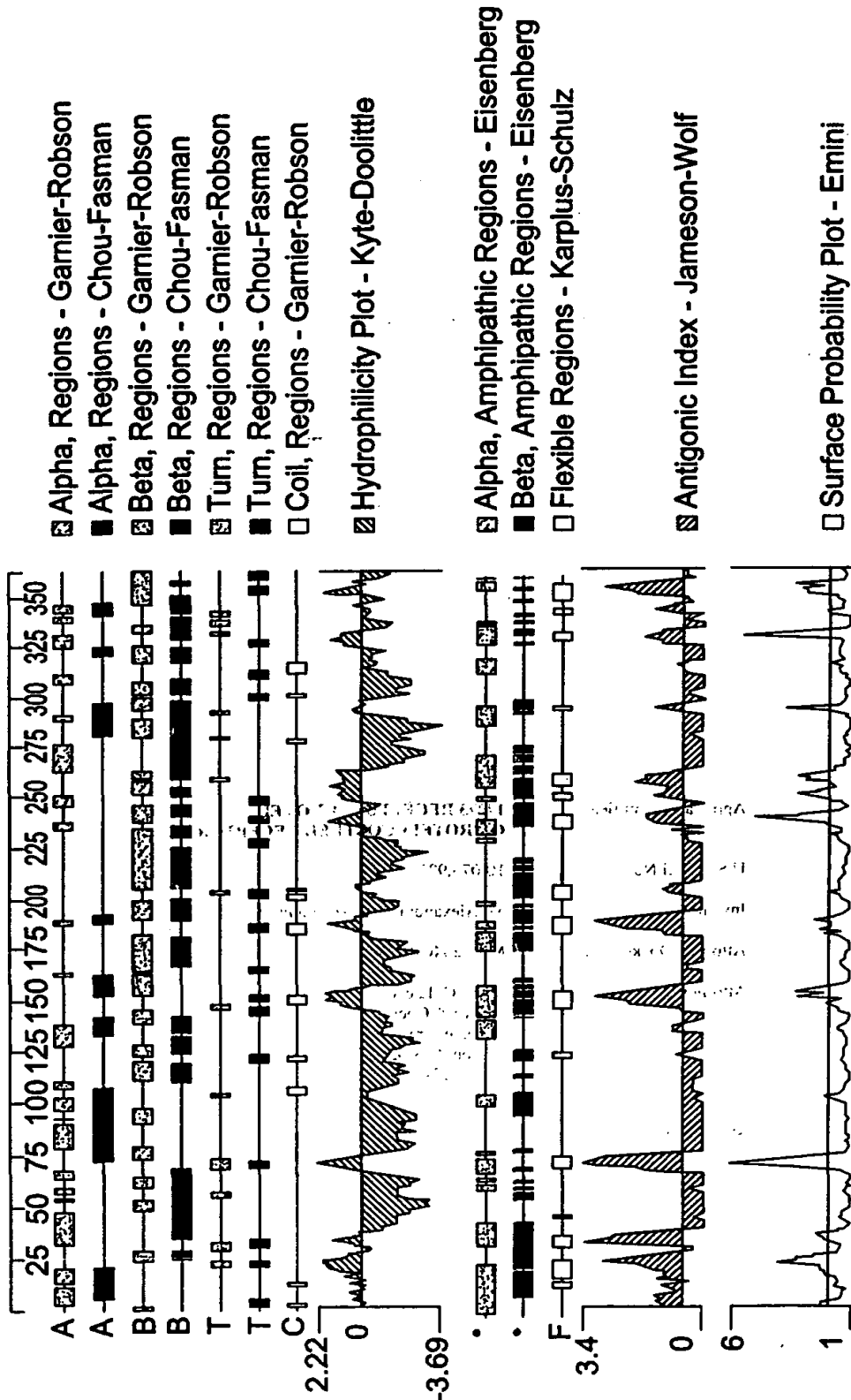


FIG. 3.

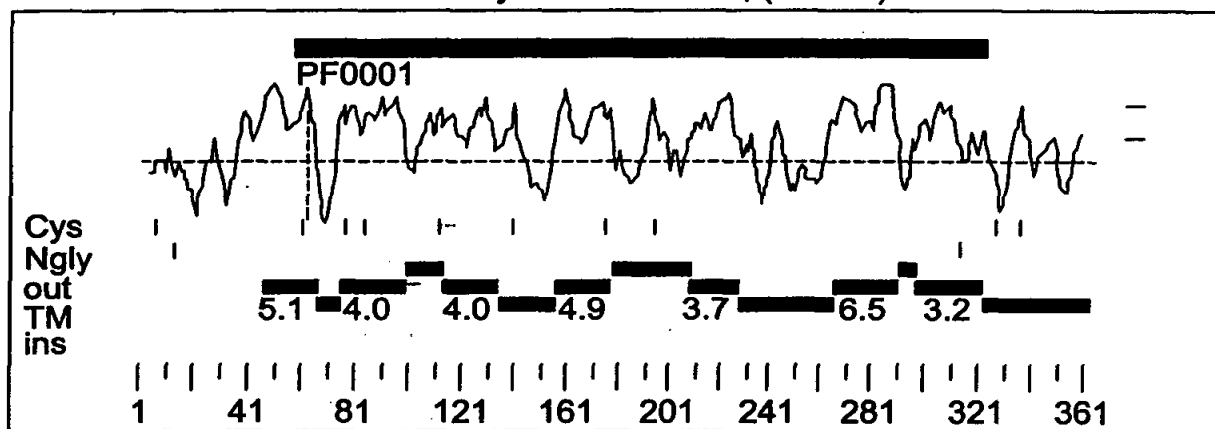


10075987.072202

6/15

COPY OF PAPERS  
ORIGINALLY FILED

## Analysis of Flh14273, (362 aa)



>Flh14273, 1086 bases, 1825 checksum.  
MSPECARAAQDAPLRSLAQNRTRFPFFSDVKGDHRLVLA AVETTVLVLEFAVSLLGNVC  
ALVLVARRRRRGATAQLVLNLF CADLLFYSAIPLVLAVRWTEAVLLGPVACHLLFYVMTL  
SGSVTILTLAAVSLERMVCIVHLQRCVRGPGRRARAVLLALIWGYSABAALPLCVFFRVV  
PQRLPGADQEISICTLIWPTIPGEISWDVSFVTLNFLVPGLVIVISYSKILQITKASRKR  
LTVSLAYSESHQIRVSQQDFRLFRTLFLLMVSFIMWSPITITILLILIQNFQDLVIWP  
SLFFWVVAFTFANSALNPILYNMTCRNEWKKIFCCPWFGEKGAILTDTSVKRNDLSIIS  
G+

FIG. 4.



10075987.072202

7/15

COPY OF PAPERS  
ORIGINALLY FILED

# Prositate Pattern Matches for F1h14273

>PS00001/PD0C00001/ASN\_GLYCOSYLATION N-glycosylation site.

Query: 21 NRTR 24

Query: 322 NMTL 325

>PS00004/PD0C00004/CAMP\_PHOSPHO\_SITE cAMP- and cGMP-dependent protein kinase phosphorylatic

Query: 239 KRLT 242

>PS00005/PD0C00005/PKC\_PHOSPHO\_SITE Protein kinase C phosphorylation site.

Query: 237 SRK 239

Query: 350 SVK 352

>PS00006/PD0C00006/EK2\_PHOSPHO\_SITE Casien kinase II phosphorylation site.

Query: 256 SQQD 259

>PS00008/PD0C00008/MYRISTYL N-myristoylation site.

Query: 57 GNVCAL 62

Query: 72 GATACL 77

Query: 343 GAILTD 348

>PS00009/PD0C00009/AMIDATION Amidation site.

Query: 150 PCRR 153

>PS00029/PD0C00029/LEUCINE\_ZIPPER Leucine zipper pattern.

Query: 106 LGPVACHLLFYVMTLSGSVTIL 127

FIG. 5.



10075987.072202

COPY OF PAPERS  
ORIGINALLY FILED

8/15

## Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
46	66	out->ins	5.1
75	98	ins->out	4.0
113	134	out->ins	4.0
156	177	ins->out	4.9
209	227	out->ins	3.7
266	289	ins->out	6.5
297	321	out->ins	3.2

&gt;F1h14273,

MSPECARAAGDAPLRSLAQANRTRFPFFSDVKGDHRLVLA AVETTVLVLIPAVSLLGNVC  
ALVLVARRRRRGATACLVNLFCADLLFISAWPLVLAVRWTEAWLLGPVACHLLFYVMTL  
SGSVTILTLAAVSLERMVCIVHLQRGVGRGRRARAVLLALIWGYSAAALPLCVFFRVV  
PQRLPGADQEISICTLIWPTIPGEISWDVSFVTNFLVPGLVIVISYSKILQITKASRKR  
LTVSLAYSESHQIRVSQQDFRLFRTLFLLMVSFFIMWSPIIITILLILIQNFKQDLVIWP  
SLFFWVVAFTFANSALNPILYNMTLCRNEWKKIFCCWFPEKGAILTDTSVKRNDLSIIS  
G

## Transmembrane Segments for Presumed Mature Peptide

Start	End	Orient	Score
14	37	ins->out	4.0
52	73	out->ins	4.0
95	116	ins->out	4.9
148	166	out->ins	3.7
205	228	ins->out	6.5
236	260	out->ins	3.2

&gt;F1h14273, \_mature

LVLVARRRRRGATACLVNLFCADLLFISAIPLVLAVRWTEAWLLGPVACHLLPYVMTLS  
GSVTILTLAAVSLERMVCIVHLQRGVGRGRRARAVLLALIWGYSAAALPLCVFFRVVP  
QRLPGADQEISICTLIWPTIPGEISWDVSFVTNFLVPGLVIVISYSKILQITKASRKR  
LTVSLAYSESHQIRVSQQDFRLFRTLFLLMVSFFIMWSPIIITILLILIQNFKQDLVIWPS  
LFFWVVAFTFANSALNPILYNMTLCRNEWKKIPCCWFPEKGAILTDTSVKRNDLSIISG

FIG. 6.





10075987.072202

COPY OF PAPERS  
ORIGINALLY FILED

9/15

Input file 14273mj Output File 14273mtra  
Sequence length 1560

TTGCCAAGCTCAGCGTAAGCCTCTTCCACTGCAATCTCACAGAAGGGTTTCATGGAGTGCTTCACACCATCAGTGACCA

CTCCAGACTGTCCGGCTTTACCCGAATCTTCACAGCGGAGTCGATGACCCCTCTTGACAGCCACGAGCGCGCAGCTC

	M	S	P	E	C	A	Q	T	G	
CGCCATCTCCCGACGCGTGGCGGCGCGCGGC	ATG	TCC	CCT	GAG	TGT	GCA	CAG	ACG	ACG	GGC
P G P S H T L D Q V N R T H F P F S D										30
CCT GCT CCC TCG CAC ACC CTG GAC GAA GTC AAT CGC ACC CAC TTC CCT TTC TTC TCG GAT										30
V K G D H R L V E S V V E T T V L G L I										50
GTC AAG GGC GAC CAC CGG TTG GTG AGC GTC GTG GAG ACC ACC GTT CTG GGA CTC ATC										150
F V V S L L G N V C A L V L V A R R R										70
TTT GTC GTC TCA CTG CTG GGC AAC TTG TGT GCT CTA GTG CTG GTG GCG CGC CGT CGG CGC										210
R G A S A S L V L N L F C A D L L F T S										90
CGT GGG GCG TCA GCC AGC CTG GTG CTC AAC CTC TTC TGC GCG GAT TTG CTC TTC ACC AGC										270
A I P L V L V V R W T E A W L L G P V										110
GCC ATC CCT CTA GTG CTC GTG GTG CGC TGG ACT GAG GCC TGG CTG TTG GGG CCC GTC GTC										330
C H L L F Y V M T M S G S V T I L T L A										130
TGC CAC CTG CTC TTC TAC GTG ATG ACA ATG AGC AGC AGC GTC ACG ATC CTC ACA CTG GCC										390
A V S L E R M V C I V R L R R G L S G P										150
GCG GTC AGC CTG GAG CGC ATG GTG TGC ATC GTG CGC CTC CGG CGC GGC TTG AGC GGC CCG										450
G R R T Q A A L L A F I W G Y S A L A A										170
GGG CGG CGG ACT CAG GCG GCA CTG CTG GCT TTC ATA TGG GGT TAC TCG GCG CTC GCC GCG										510

TD FIG. 7B.

FIG. 7A.



10075987.072202

COPY OF PAPERS  
ORIGINALLY FILED

10/15

FROM FIG. 7A.

L P L Y I L L F R V V P Q R L P G G D Q E 190  
CTG CCC CTC TAC ATC TTG TTC CGC GTG GTC CCG CAG CGC CTT CCC GGC GGC GAC CAG GAA 627  
I P I C T L L D W P N R I G E I S W D V F 210  
ATT CCG ATT TGC ACA TTG GAT TGG CCC AAC CGC ATA GGA GAA ATC TCA TGG GAT GTG TTT 630  
F E T L N F L V P G L V I V I S Y S K I 230  
TTT GAG ACT TTG AAC TTC CTG GTG CCG GGA CTG GTG ATT GTG ATC AGT TAC TCC AAA ATT 690  
L Q I T K A S TCG CGG AAG AGG CTT ACG CTG AGC TTG GCA TAC TCT GAG AGC 250  
TTA CAG ATC ACG AAA GCA TCG CGG AAG AGG CTT ACG CTG AGC TTG GCA TAC TCT GAG AGC 750  
H Q I R V S Q Q Q R I I I T I L L F L L M 270  
CAC CAG ATC CGA GTG TCC CAA CAA CAC TAC CGA CTC TTC CGC AGC CTC TTC CTG CTC ATC 810  
V S F F I M W S P I I I I L L I I Q 290  
GTT TCC TTC TTC ATC ATG TGG AGT TCC ATC ATC ACC ATC CTC CTC ATC TTC ATC CAA 870  
N F R Q D L V I W P S L F F W V A F T 310  
AAC TTC CGG CAG GAC CTG GTG ATC TGG CCA TCC CTT TTC TTC TGG GTG GTG GCC TTC ACG 930  
F A N S A L N P I L Y N M S L F R N E W 330  
TTT GCC AAC TCT GCC CTA AAC CCC ATA CTG TAC AAC ATG TCG CTG TTC AGG AAC GAA TGG 990  
R K I F C C F F F P E K G A I F T D T S 350  
AGG AAG ATT TTT TGC TGC TTC TTT TTT TTT CCA GAG AAG GGA GCC ATT TTT ACA GAT ACG TCT 1050  
V R R N D L S V I S S • 362  
GTC AGG CGA AAT GAC TTG TCT GTT ATT TCC AGC TAA 1086

CTAGCCTCTGGTCCAGGTGAACCACGGTGTGTCATGTAAAGCGAGTTAACTTCAAGGAAGCCACCAGTGCCTGCTGC  
TTTAAATACCGACTTCCACAGCAGGCATCTACGGAGCCAGCAAAATTAAAGGAATGATCGCTCAGTATAAAATATT  
TTTCCTTAAAGAACCTTCTATGGTTCCTTTTGTGAACCTTTTAAAGTGTTTGTAAATATGATCTAGTTAATAATT  
TTTATTATAACGTGTTCTACAAAAAATAAAAAAAAAAAAAA

FIG. 7B.

**COPY OF PAPERS  
ORIGINALLY FILED**

<u>Sequence</u>	<u>Description</u>	<u>Score</u>	<u>E-value</u>	<u>N</u>
7tm_1	<u>PF00001</u> 7 transmembrane receptor (rhodopsin	118.8	1e-35	1

Sequence	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value	
7tm_1	1/1	57	321	..	1	259 [ ]	118.8	1e-36

14273m,290 QnFR-----QDLVIWPSLFFWVVAFTFANSALNPILY 321

**FIG. 8.**



10075087.072202

COPY OF PAPERS  
ORIGINALLY FILED

12/15

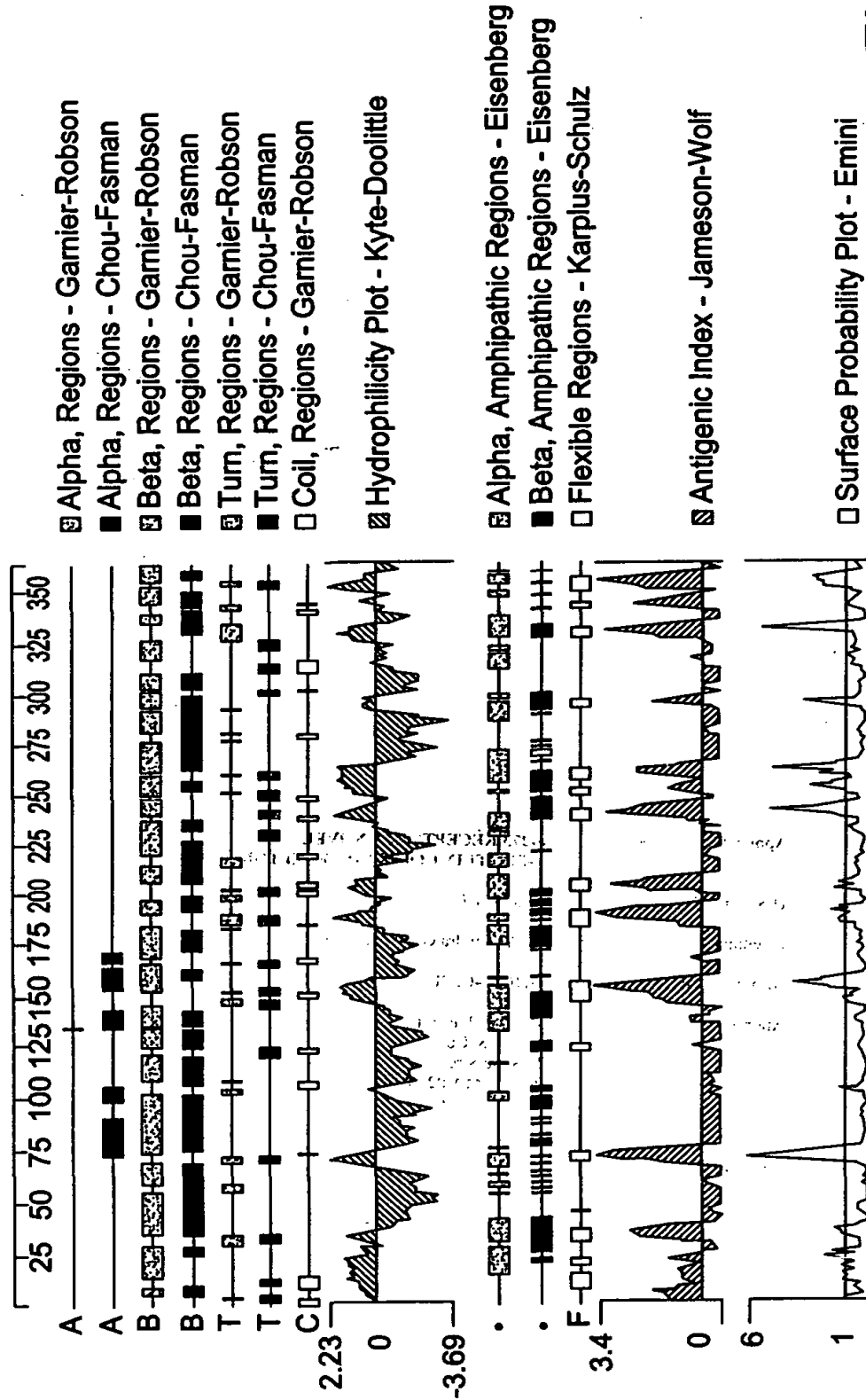


FIG. 9.

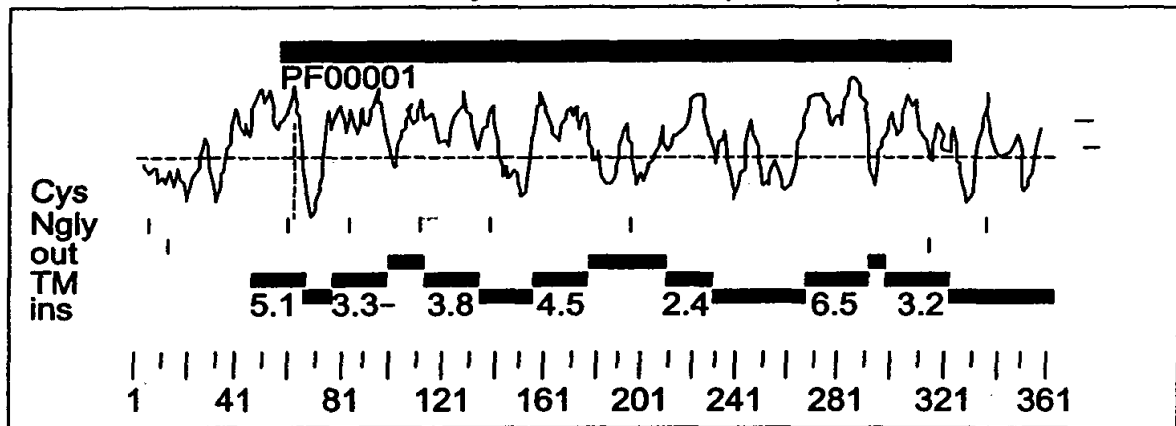


10075987.072202

13/15

COPY OF PAPERS  
ORIGINALLY FILED

## Analysis of 14273m, (362 aa)



>14273m, 1086 bases, 6943 checksum.  
MSPECAQT TGPCSH TLDQVNRTHFPFFSDVKGDHRLVLSVETT VLGLIFVVSLLGNVC  
ALVLVARRRRRGASASLVNLF CADLLF TSAIPLVLVVRWTEAWLLGPVCHLLFVMTM  
SGSVTIL TLAAYSLERMVCIVRLRRGLSGPGRRTQAALLAFI WGYSALAALPLYILFRVV  
PQRLPGGDQEIPIC TLDWPNRIGEISWDVFFETLNLV PGLVIVISYSKILQITKASRR  
LTLSLAYSESHQIRVSQQDYRLFRTLFLMVSEFFI WSP IITILLIQNFRQDLVIWP  
SLFFVWVAFTFANSALNPILYNMSLFRNEWRKIFCCFPPEKGAIFTDTSVRRNDLSVIS  
S\*

FIG. 10.



10075987.072202

14/15

COPY OF PAPERS  
ORIGINALLY FILED

Prosite Pattern Matches for 14273m,

&gt;PS00001/PDOC00001;ASN\_GLYCOSYLATION N-glycosylation site.

Query: 21	NRTH	24
Query: 322	NRTH	325

&gt;PS00002/PDOC00002/GLYCOSAMINOGLYCAN Glycosaminoglycan attachment site.

RU	Additional rules:
RU	There must be at least two acidic amino acids (Glu or Asp) from -2 to -4 relative to the serine.

Query: 148	SGPG	151
------------	------	-----

&gt;PS00004/PDOC00004/CAMP\_PHOSPHO\_SITE cAMP- and cGMP-dependent protein kinase phosphorylation

Query: 239	KRLT	242
------------	------	-----

&gt;PS00005/PDOC00005/PKC\_PHOSPHO\_SITE Protein kinase C phosphorylation site.

Query: 237	SRK	239
Query: 350	SVR	352

&gt;PS00006/PDOC00006/CK2\_PHOSPHO\_SITE Casein kinase II phosphorylation site.

Query: 40	SVVE	43
Query: 256	SQQD	259

&gt;PS00008/PDOC00008/MYRISTYL N-myristoylation site.

Query: 57	GNVCAL	62
Query: 72	GASASL	77
Query: 343	GAIFTD	348

&gt;PS00009/PDOC00009/AMIDATION Amidation site.

Query: 150	PGRR	153
------------	------	-----

FIG. 11.



10075087.072202

15/15

COPY OF PAPERS  
ORIGINALLY FILED

## Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
46	66	out->ins	5.1
77	98	ins->out	3.3
113	134	out->ins	3.8
156	177	ins->out	4.5
209	227	out->ins	2.4
266	289	ins->out	6.5
297	321	out->ins	3.2

&gt;14273m,

MSPECAQTTGPGPSHTLDQVNRTHFPFFSDVKGDHRLVLSVVETTVLGLIFVVSLLGNVC  
ALVLVARRRRRGASASLVNLFCADLLFTSAIPLVLVVRWTEAWLLGPVVCHLLFYVMTM  
SGSVTILTLAAVSLERMVCIVRLRRLSGPGRRTQAALLAFIWGYSALAALPLYILFRVV  
PQRLPGGDQEIPICITLDWPNRIGESISWDVFFETLNFLVPGLVIVISYSKILQITKASRKR  
LTLSEYSESHQIRVSQQDYRLFRTLFLLMVSFFIMWSPIIITILLILIQNFRQDLVIWP  
SLFFWVVAFTFANSALNPILYMMSLFRNEWKIFCCFFFPEKGAIFTDTSVRRNDLSVIS  
S

## Transmembrane Segments for Presumed Mature Peptide

Start	End	Orient	Score
16	37	ins->out	3.3
52	73	out->ins	3.8
95	116	ins->out	4.5
148	166	out->ins	2.4
205	228	ins->out	6.5
236	260	out->ins	3.2

&gt;14273m, \_mature

LVLVARRRRRGASASLVNLFCADLLFTSAIPLVLVVRWTEAWLLGPVVCHLLFYVMGMS  
GSVTILTLAAVSLERMVCIVRLRRLSGPGRRTQAIIAFIWGYSALAALPLYILFRVVP  
QRLPGGDQEIPICITLDWPNRIGESISWCVFFETLNFLVPGLVIVISYSKILQITKASRKR  
LTLSEYSESKQIRVSQQDYRLFRTLFLLMVSFFTMWSPIIITILLILIQNFRQDLVIWPS  
LFFWVVAFTFANSALNPILYNMSLFRNEWKIFCCFFFPEKGAIFTDTSVRRNDLSVISS

**FIG. 12.**